①

AD-A223 832

RECENT ADVANCES IN STATISTICAL RANKING AND SELECTION: THEORY AND METHODS *

by

Shanti S. Gupta
Purdue University

Technical Report #90-27C

DTIC
ELECTE
JUL 11 1990
S   D
D^{α}

Department of Statistics
Purdue University

June 1990

*

90 07 9 052

Recent Advances in Statistical Ranking
and Selection: Theory and Methods *

by

Shanti S. Gupta
Department of Statistics
Purdue University

Abstract

Ranking and selection procedures have been developed in modern statistical methodology over the past 35 years with fundamental papers pioneered by Bechhofer (1954) and Gupta (1956). Since then, various modifications and applications have taken place. The reader is referred to Gupta and Panchapakesan (1979, 1985) for an overview of this research area.

In the present paper, we provide a review of some recent developments in the research area of ranking and selection, mainly, on the following topics: (A) Selecting the largest normal mean and estimating the selected mean, (B) empirical Bayes selection, (C) selecting the important regression variables, (D) sequential selection rules, and (E) lower confidence bounds for the probability of a correct selection. Related theoretic and methodological research will be surveyed. Aspects of ongoing research will also be discussed.

AMS Subject Classification: 62F07, 6202; 62C12, 62J05

Key Words: Ranking and Selection, Empirical Bayes Procedures, Asymptotic Optimality, Probability of Correct Selection, Regression Variables, Sequential Selection Procedures.

1. Introduction

In many practical situations, the goal of the experimenter is to compare two or more populations in order to make a decision in the form of a ranking of the populations. The best studied ranking goal concerns the best population (the most efficient drug for an ailment, the most effective manufacturing process and so on). The classical tests of homogeniety were not designed to provide answers to such questions. Rejecting the null hypothesis is not the final solution to the experimenter's problem but an exercise that underlies the need for a reformulation of the problem. Born out of this need is the statistical theory of ranking and selection procedures.

Ranking and selection problems have been generally formulated using either the *indifference zone approach* due to Bechhofer (1954) or the *subset selection approach* of Gupta (1956, 1965). Starting from the early developments in the 1950's, these problems have been extensively studied under various model assumptions and modifications in the ranking goals. A comprehensive survey of these developments is provided in Gupta and Panchapakesan (1979), who have in a later paper (1985) given a review of these and subsequent developments with historical perspectives.

In the present paper, we review some recent developments in the ranking and selection theory. We will focus our attention on the following topics: (A) Selecting the largest normal mean and estimating the selected mean, (B) Empirical Bayes selection, (C) Selecting the important regression variables, (D) Sequential selection rules, and (E) Lower confidence bounds for the probability of a correct selection.

2. Finding the Largest Normal Mean and Estimating the Selected Mean

Let $\pi_1, \ldots, \pi_k$ be $k(\geq 2)$ normal populations with unknown means $\theta_1, \ldots, \theta_k$ and a common known variance σ^2. Let $\theta_{[1]} \leq \ldots \leq \theta_{[k]}$ denote the ordered θ_i. The population associated with $\theta_{[k]}$ is called the best population. The goal is to select one of the k populations as the best. Since no procedure assures the selection of the best with certainty, estimation of the mean of the selected population is of practical interest.

Let X_i denote the sample mean of n_i independent observations from π_i, $i = 1, \ldots, k$. The so-called *natural selection rule* selects as the best the population that yields the largest X_i. When the sample sizes $n_1, \ldots, n_k$ are all equal, this rule is the uniformly best permutation invariant selection rule for a general class of loss functions. However, for unequal sample sizes, the natural selection rule loses much of its optimality (see Gupta and Miescke (1988)).

A natural estimator of $\theta_{(k)}$, the selected θ_i, is $X_{[k]}$, the largest X_i. However, $X_{[k]}$ overestimates $\theta_{[k]}$ and thus overestimates $\theta_{(k)}$ even more. Recognizing this, alternative estimators have been studied for the present and other experimental models and goals by the following authors: Sarkadi (1967), Dahiya (1974), Cohen and Sackrowitz (1982), Sackrowitz and Samuel-Cahn (1984,1986), Jeyaratnam and Panchapakesan (1984,1986,1988), Vellaisamy and Sharma (1988,1989), Vellaisamy, Kumar and Sharma (1988), and Venter (1988). Since selection is made first, the preceding estimation problem is known as *estimation after selection.*

Cohen and Sackrowitz (1988) presented a decision-theoretic framework for the combined selection-estimation problem, and derived results for the case of $k = 2$ and $n_1 = n_2$. Recently, Gupta and Miescke (1990) have extended the results of Cohen and Sackrowitz (1988) and provided a detailed discussion for normal distributions problem. Rather than the "estimating after selection," the decision-theoretic treatment of the combined decision problem leads to "selecting after estimation."

We will first discuss the decision-theoretic approach under a general framework and then examine the normal means case.

2.1 General Framework

Let $X = (X_1, \ldots, X_k)$ be a random vector of observations having pdf $f(x|\theta) = \prod_{i=1}^{k} f_i(x_i|\theta_i)$, where $x = (x_1, \ldots, x_k)$ and $\theta = (\theta_1, \ldots, \theta_k)$. Here, X may be a vector of sufficient statistics for $\theta_1, \ldots, \theta_k$. The goal is to select the "population" associated with $\theta_{[k]}$ and to simultaneously estimate θ_s, the selected θ-value. For this combined problem,

3

a nonrandomized decision rule is:

$$\underline{d}(\underline{x}) = (s(\underline{x}), \ \ell_{s(\underline{x})}(\underline{x})) \tag{2.1}$$

where $s(\underline{x}) \in \{1, 2, \ldots, k\}$ is the selection rule, and $\ell_i(\underline{x})$ is an estimate for $\theta_i, i = 1, \ldots, k$. We assume an additive loss function $L(\underline{\theta}, \underline{d})$ given by

$$L(\underline{\theta}, \underline{d}) = A(\underline{\theta}, s) + B(\theta_s, \ell_s) \tag{2.2}$$

where A is the loss incurred in selecting π_s as the best population when $\underline{\theta}$ is the true parametric vector, and B is the loss of estimating θ_s by ℓ_s.

Adopting the Bayes approach, we assume that $\underline{\Theta} = (\Theta_1, \ldots, \Theta_k)$ has a prior distribution G. Then, for $\underline{X} = \underline{x}$, the posterior risk of $\underline{d}(\underline{x})$ can be expressed as:

$$r(\underline{d}(\underline{x})) = r_A(s(\underline{x})) + r_B(s(\underline{x}), \ \ell_{s(\underline{x})}(\underline{x})), \tag{2.3}$$

where
$$r_A(s(\underline{x})) = E\{A(\underline{\Theta}, s(\underline{x})) | \underline{X} = \underline{x}\}, \text{ and}$$

$$r_B(s(\underline{x}), \ell_{s(\underline{x})}(\underline{x})) = E\{B(\Theta_{s(\underline{x})}, \ell_{s(\underline{x})}(\underline{x})) | \underline{X} = \underline{x}\}.$$

The following theorem of Gupta and Miescke (1990) is an extension of a result of Cohen and Sackrowitz (1988).

Theorem 2.1. Let $\ell_i^*(\underline{x})$ minimize $r_B(i, \ell_i(\underline{x}))$, $i = 1, \ldots, k$, and let $s^*(\underline{x})$ minimize $r_A(s(\underline{x})) + r_B(s(\underline{x}), \ell_{s(\underline{x})}^*(\underline{x}))$. Then the Bayes decision rule is:

$$\underline{d}^*(\underline{x}) = (s^*(\underline{x}), \ \ell_{s^*(\underline{x})}^*(\underline{x})).$$

Remark 2.1. It can be seen that the combined selection–estimation problem is in a sense "selecting after estimation."

Corollary 2.1. Whenever at $\underline{X} = \underline{x}$, $r_B(i, \ell_i^*(\underline{x}))$ does not depend on $i \in \{1, 2, \ldots, k\}$, $s^*(\underline{x})$ minimizes $r_A(s(\underline{x}))$.

Let $s^N(\underline{x})$ denote the natural selection rule which selects the population corresponding to the largest x_i. The following example shows that s^N is *not* same as s^* in general.

Example 2.1. Let $X_i \sim N(\theta_i, 1)$, $i = 1, \ldots, k$, be independent. Assume that $\Theta_1, \ldots, \Theta_k$ are iid having the density $\exp(-\theta)$, $\theta > 0$, and consider a loss function given by $L(\underset{\sim}{\theta}, \underset{\sim}{d}) = A(\underset{\sim}{\theta}, s) + (\theta_s - \ell_s)^2$, where A is permutation invariant and favors selecting large θ–values. A posteriori, at $\underset{\sim}{X} = \underset{\sim}{x}$, $\Theta_1, \ldots, \Theta_k$ are independent and the posterior density of Θ_i is $\varphi(\theta_i - y_i)/\Phi(y_i)$, $\theta_i > 0$, where φ and Φ denote the $N(0,1)$ density and cdf, respectively, and $y_i = x_i - 1, i = 1, \ldots, k$. Straightforward computations yield, for $i = 1, \ldots, k$,

$$
\begin{cases}
\ell_i^*(\underset{\sim}{x}) & = E[\Theta_i | \underset{\sim}{X} = \underset{\sim}{x}] = y_i + \frac{\varphi_i(y_i)}{\Phi(y_i)} \\
r_B(i, \ell_i^*(\underset{\sim}{x})) & = \mathrm{Var}[\Theta_i | \underset{\sim}{X} = \underset{\sim}{x}] \\
& = 1 + 2y_i^2 + \frac{y_i \varphi(y_i)}{\Phi(y_i)} - \left[\frac{\varphi(y_i)}{\Phi(y_i)}\right]^2.
\end{cases}
$$

Thus, although $s^N(\underset{\sim}{x})$ minimizes $r_A(s(\underset{\sim}{x})), s^N$ is different from s^*, since $r_B(i, \ell_i^*(\underset{\sim}{x}))$ depends on $i \in \{1, \ldots, k\}$.

2.2. Normal Means Problem

Let X_i denote the mean of a random sample of size n_i from $N(\theta_i, \sigma^2)$ population, $i = 1, \ldots, k$, where the common variance σ^2 is assumed to be known. Apriori, $\Theta_1, \ldots, \Theta_k$ are independent and $\Theta_i \sim N(\mu_i, q_i)$, $i = 1, \ldots, k$. Thus, given $\underset{\sim}{X} = \underset{\sim}{x}$, $\Theta_1, \ldots, \Theta_k$ are aposteriori independent with $\Theta_i | \underset{\sim}{X} = \underset{\sim}{x} \sim N\left(\frac{q_i x_i + p_i \mu_i}{q_i + p_i}, \frac{p_i q_i}{p_i + q_i}\right)$, $i = 1, \ldots, k$. Also, the X_i's are marginally independent with $X_i \sim N(\mu_i, p_i + q_i)$, where $p_i = \sigma^2/n_i$, $i = 1, \ldots, k$.

Equal Sample Sizes. We let $n_1 = \ldots = n_k = n$. We also assume that $\mu_1 = \ldots = \mu_k = \mu$ and $q_1 = \ldots = q_k = q$, i.e. we have exchangeable normal priors. We assume the loss function $L(\underset{\sim}{\theta}, \underset{\sim}{d})$ in (2.2) with two possible forms $B_1(\theta_s, \ell_s)$ and $B_2(\theta_s, \ell_s)$ for $B(\theta_s, \ell_s)$, given by

$$
\begin{cases}
B_1(\theta_s, \ell_s) & = |\theta_s - \ell_s| \\
B_2(\theta_s, \ell_s) & = (\theta_s - \ell_s)^2.
\end{cases}
\tag{2.4}
$$

Also, $A(\underset{\sim}{\theta}, s)$ in (2.2) is assumed to be permutation symmetric and favorable to selection of larger θ–values.

Under the above assumptions, Gupta and Miescke (1990) have shown that the Bayes rule $\underset{\sim}{d}^* = (s^*, \ell_s^*.)$ satisfies $s^* = s^N$ and $\ell_i^*(\underset{\sim}{x}) = E\{\Theta_i | \underset{\sim}{X} = \underset{\sim}{x}\}$, $i = 1, \ldots, k$.

Consider the natural decision rule $\underset{\sim}{d}^N = (s^N, \ell_{s^N}^N)$, where $\ell_i^N(\underset{\sim}{x}) = x_i$, $i = 1, \ldots, k$. Although, from the frequentist point of view, $\underset{\sim}{d}^N$ has the undesirable feature of overes-

5

timating the mean of the selected population, it has been shown by Gupta and Miescke (1990) to be an extended Bayes rule.

Unequal Sample Sizes Case. Here we will consider two particular loss functions, namely,

$$\begin{cases} L_1(\underline{\theta}, \underline{d}) & = c(\theta_{[k]} - \theta_s) + |\theta_s - \ell_s|, \\ L_2(\underline{\theta}, \underline{d}) & = c(\theta_{[k]} - \theta_s)^2 + (\theta_s - \ell_s)^2, \end{cases} \quad (2.5)$$

where $c > 0$ gives relative weights to the two parts of the loss function. Since the sample sizes are unequal, it is appropriate to consider non–exchangeable priors, $\Theta_i \sim N(\mu_i, q_i)$, $i = 1, \ldots, k$. The Θ_i's are, of course, independent.

Under the above setup, *with loss function L_1*, the Bayes rule (by Theorem 2.1) employs the estimator $\ell_i^*(\underline{x}) = (q_i x_i + p_i \mu_i)/(q_i + p_i)$ for θ_i, $i = 1, \ldots, k$, and one has to find $s^*(\underline{x})$. For any decision rule $\underline{d} = (s, \ell_s^*)$, the posterior risk at $\underline{X} = \underline{x}$ associated with selection $s(\underline{x}) = i \in \{1, \ldots, k\}$ turns out to be

$$c \left[E\{\Theta_{[k]} | \underline{X} = \underline{x}\} - \frac{q_i x_i + p_i \mu_i}{q_i + p_i} \right] + \left(\frac{2 p_i q_i}{\pi(q_i + p_i)} \right)^{\frac{1}{2}}.$$

This leads to the Bayes rule $\underline{d}^* = (s^*, \ell_s^*)$, where $\ell_i^*(\underline{x}) = (q_i x_i + p_i \mu_i)/(q_i + p_i)$, $i = 1, \ldots, k$, and $s^*(\underline{x})$ maximizes $c\ell_i^*(\underline{x}) - [2 q_i p_i / \pi(q_i + p_i)]^{\frac{1}{2}}$, $i = 1, \ldots, k$.

It is interesting to note three special cases, which are as follows.

Case 1: Noninformative prior ($q_i \to \infty$, $i = 1, \ldots, k$). In this case, $\ell_i^*(\underline{x}) = x_i$, $i = 1, \ldots, k$, and $s^*(\underline{x})$ maximizes $x_i - c^{-1}(2 p_i / \pi)^{\frac{1}{2}}$, $i = 1, \ldots, k$.

Case 2: Prior variances proportional to sample variances ($q_i = \gamma p_i$, $i = 1, \ldots, k$, for some $\gamma > 0$). In this case, $\ell_i^*(\underline{x}) = (\gamma x_i + \mu_i)/(\gamma + 1)$, $i = 1, \ldots, k$, and $s^*(\underline{x})$ maximizes $\ell_i^*(\underline{x}) - c^{-1}\{2 \gamma p_i / (\gamma + 1)\pi\}^{\frac{1}{2}}$, $i = 1, \ldots, k$. In particular, for $\mu_1 = \ldots = \mu_k = \mu$ (say), $\ell_i^*(\underline{x}) = (\gamma x_i + \mu)/(\gamma + 1)$, $i = 1, \ldots, k$, and $s^*(\underline{x})$ maximizes $x_i - c^{-1}\{2(\gamma + 1)p_i / \pi\}^{\frac{1}{2}}$, $i = 1, \ldots, k$.

Case 3: Posterior decreasing in transposition (DT), i.e. $q_i^{-1} + p_i^{-1} = \tau^{-1}$, $i = 1, \ldots, k$, for some fixed $\tau > 0$. In this case, $\ell_i^*(\underline{x}) = \tau(p_i^{-1} x_i + q_i^{-1} \mu_i)$, $i = 1, \ldots, k$, and $s^*(\underline{x})$ maximizes $\ell_i^*(\underline{x})$, $i = 1, \ldots, k$. In particular, when $\mu_1 = \ldots = \mu_k = \mu$ (say), $\ell_i^*(\underline{x}) = p_i^{-1} \tau(x_i - \mu) + \mu$, $i = 1, \ldots, k$, and $s^*(\underline{x})$ maximizes $p_i^{-1}(x_i - \mu)$, $i = 1, \ldots, k$.

It has been shown by Gupta and Miescke (1990) that the decision rule of Case 1 (noninformative prior) is an extended Bayes rule.

In the case of loss function L_2, the analysis gets more complicated. For finding the Bayes rule, we have the same $\ell_i^*(x)$ as in the case of L_1, but for finding $s^*(x)$ one has to minimize

$$cE\left\{[\Theta_{[k]} - \Theta_i]^2 | X = x\right\} + \frac{q_i p_i}{q_i + p_i}. \qquad (2.6)$$

The difficulty lies in the fact that, for any i, the conditional distribution of $(\Theta_{[k]}, \Theta_i)$ at $X = x$ does not yield simpler representations for the conditional expectation in (2.6), which in most situations has to be evaluated on a computer.

As in the case of L_1, we can specialize the problem in the three special cases regarding the assumptions about the prior. In Case 3, the Bayes rule is the same as in the case of L_1.

3. Empirical Bayes Selection Procedures

The empirical Bayes approach in statistical decision theory is typically appropriate when one is confronted repeatedly and independently with the same decision problem. In such instances, it is reasonable to formulate the component problem with respect to an unknown prior distribution on the parameter space. One then uses information borrowed from other sources to improve the decision procedure for each component. This approach is due to Robbins (1956,1964). Empirical Bayes procedures have been derived for multiple decision problems by Deely (1965). Recently, Gupta and Hsiao (1983), Gupta and Leu (1983), and Gupta and Liang (1986, 1988a,b, 1989a,b,c) have investigated empirical Bayes procedures for several selection problems. Many such empirical Bayes procedures have been shown to be asymptotically optimal in the sense that the component Bayes risk will converge to the optimal Bayes risk which would have been obtained if the prior distribution were fully known, and the Bayes procedure with respect to this prior distribution was used.

In this section, we will describe empirical Bayes selection procedures with respect to a standard. Two kinds of empirical Bayes procedures will be considered. One is to incor-

7

porate past data to improve the current decision. The other is to incorporate information from each other so as to simultaneously improve the decisions for each of the component problems under study. A Poisson distribution model is used as an example to describe the empirical Bayes idea and methods.

3.1. Formulation of the Empirical Bayes Selection Problem

Let $\pi_1, \ldots, \pi_k$ denote k independent populations. For each $i = 1, \ldots, k$, let X_i denote a random observation arising from population π_i. It is assumed that X_i follows a Poisson distribution with probability function $f_i(x|\theta_i)$ where

$$f_i(x|\theta_i) = e^{-\theta_i}\theta_i^x/x!, \ x = 0, 1, 2, \ldots; \ \theta_i > 0.$$

Let $\theta_0 > 0$ be a known standard. Population π_i is said to be *good* if $\theta_i \geq \theta_0$, and *bad* otherwise. The goal is to select all the good populations and exclude all the bad populations.

Let $\Omega = \{\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_k)|\theta_i > 0, \ i = 1, \ldots, k\}$ be the parameter space and let $A = \{\underset{\sim}{a} = (a_1, \ldots, a_k)|a_i = 0, 1, \ i = 1, \ldots, k\}$ be the action space. When action $\underset{\sim}{a}$ is taken, it means that population π_i is selected as a good population if $a_i = 1$, and excluded as a bad one if $a_i = 0$. For each $\underset{\sim}{\theta} \in \Omega$ and $\underset{\sim}{a} \in A$, the loss function $L(\underset{\sim}{\theta}, \underset{\sim}{a})$ is defined to be:

$$L(\underset{\sim}{\theta}, \underset{\sim}{a}) = \sum_{i=1}^{k} a_i(\theta_0 - \theta_i)I(\theta_0 - \theta_i) + \sum_{i=1}^{k}(1 - a_i)(\theta_i - \theta_0)I(\theta_i - \theta_0) \qquad (3.1)$$

where $I(x) = 1(0)$ if $x \geq (<)0$.

It is assumed that for each i, the parameter θ_i is a realization of a random variable Θ_i which has a prior distribution G_i. It is also assumed that $\Theta_1, \ldots, \Theta_k$ are mutually independent. Thus $\underset{\sim}{\Theta} = (\Theta_1, \ldots, \Theta_k)$ has a joint prior distribution $\underset{\sim}{G}(\underset{\sim}{\theta}) = \prod_{i=1}^{k} G_i(\theta_i)$.

For each $i = 1, \ldots, k$, let $\mathcal{X}_i$ be the sample space of X_i, and let $\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_k$. Let $\underset{\sim}{X} = (X_1, \ldots, X_k)$. A selection rule $\underset{\sim}{d} = (d_1, \ldots, d_k)$ is defined to be a mapping from $\mathcal{X}$ into $[0, 1]^k$ such that $d_i(\underset{\sim}{x})$ is the probability of selecting population π_i as a good population when $\underset{\sim}{X} = \underset{\sim}{x}$ is observed. Let D be the class of all selection rules, and let $r(\underset{\sim}{G}, \underset{\sim}{d})$ denote

8

the Bayes risk associated with each $\underset{\sim}{d} \in D$. Then, $r(\underset{\sim}{G}) = \underset{\underset{\sim}{d} \in D}{\inf} r(\underset{\sim}{G}, \underset{\sim}{d})$ is the minimum Bayes risk.

The Bayes risk associated with any rule $\underset{\sim}{d} \in D$ can be written as:

$$r(\underset{\sim}{G}, \underset{\sim}{d}) = \sum_{i=1}^{k} r_i(\underset{\sim}{G}, d_i) \tag{3.2}$$

where

$$r_i(\underset{\sim}{G}, d_i) = \sum_{\underset{\sim}{x} \in \mathcal{X}} [\theta_0 - \varphi_i(x_i)] d_i(\underset{\sim}{x}) \prod_{j=1}^{k} f_j(x_j) + C_i, \tag{3.3}$$

where $\varphi_i(x_i) = E[\Theta_i | X_i = x_i] = h_i(x_i + 1)/h_i(x_i)$ is the posterior mean of Θ_i given $X_i = x_i$, $h_i(x_i) = f_i(x_i)/a(x_i)$, $f_i(x_i) = \int_0^\infty f_i(x_i|\theta) dG_i(\theta) = \int_0^\infty e^{-\theta} \theta^{x_i}/x_i! dG_i(\theta) = a(x_i) h_i(x_i)$ is the marginal probability function of the random variable X_i, and $a(x_i) = (x_i!)^{-1}$, $h_i(x_i) = \int_0^\infty e^{-\theta} \theta^{x_i} dG_i(\theta)$ and $C_i = \int_{\theta_0}^\infty (\theta - \theta_0) dG_i(\theta)$.

It follows that a Bayes rule, say $\underset{\sim}{d}_B = (d_{1B}, \ldots, d_{kB})$, is clearly given by: For each $i = 1, \ldots, k$,

$$d_{iB}(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \varphi_i(x_i) \geq \theta_0; \\ 0 & \text{otherwise.} \end{cases} \tag{3.4}$$

The minimum Bayes risk is: $r(\underset{\sim}{G}) = \sum_{i=1}^{k} r_i(\underset{\sim}{G}, d_{iB})$.

When the prior distribution $\underset{\sim}{G}$ is unknown, it is not possible to apply the Bayes rule $\underset{\sim}{d}_B$ for the selection problem. In the following, the empirical Bayes approach of Robbins (1956,1964) is employed. First, we discuss the case where certain past observations from each of the k populations are available.

3.2. Incorporating Information from Past Observations

According to the usual empirical Bayes framework, it is assumed that for each $i = 1, \ldots, k$, there are marginally iid past random observations $X_{i1}, \ldots, X_{in}$ with marginal probability function $f_i(x)$ available when a decision is made. Three empirical Bayes selection rules are constructed according to how much we know about the prior distribution $\underset{\sim}{G}$.

3.2.1. A Nonparametric Empirical Bayes Rule

It is assumed that the prior distribution G is completely unknown. Thus, a nonparametric empirical Bayes approach is employed. It should be noted that $\varphi_i(x_i)$ is increasing in x_i for each $i = 1, \ldots, k$. Therefore the Bayes rule d_B is a monotone selection rule. Thus, it is desirable that the considered empirical Bayes rule be also monotone.

For each $i = 1, \ldots, k$, and $x = 0, 1, 2, \ldots$, define

$$f_{in}(x) = n^{-1} \sum_{j=1}^{n} I_{\{x\}}(X_{ij})$$
$$h_{in}(x) = f_{in}(x)/a(x).$$

Let $N_{in} = \max_{1 \leq j \leq n} X_{ij} - 1$ and for each $x = 0, 1, \ldots, N_{in}$, define

$$\varphi_{in}(x) = [h_{in}(x+1) + \delta_n]/[h_{in}(x) + \delta_n],$$

where $\delta_n > 0$ is such that $\delta_n = o(1)$.

Since $\varphi_{in}(x)$ may not be increasing in x, a smoothed version of $\varphi_{in}(x)$ is given below. Let $\{\varphi_{in}^*(x)\}_{x=0}^{N_{in}}$ be the isotonic regression of $\{\varphi_{in}(x)\}_{x=0}^{N_{in}}$ with random weights $\{W_{in}(x)\}_{x=0}^{N_{in}}$, where $W_{in}(x) = [h_{in}(x) + \delta_n]a(x+1)$. For $y > N_{in}$, let $\varphi_{in}^*(y) = \varphi_{in}^*(N_{in})$. Therefore, $\varphi_{in}^*(x)$ is nondecreasing in x. We may use $\varphi_{in}^*(x)$ to estimate $\varphi_i(x)$. Based on $\varphi_{in}^*(x)$, $i = 1, \ldots, k$, an empirical Bayes rule $d_n^* = (d_{in}^*, \ldots, d_{kn}^*)$ is proposed as follows: For each $i = 1, \ldots, k$, and $x \in \mathcal{X}$,

$$d_{in}^*(x) = \begin{cases} 1 & \text{if } \varphi_{in}^*(x_i) \geq \theta_0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.5}$$

3.2.2. A Parametric Empirical Bayes Rule

It is assumed that the prior distribution G_i is the gamma distribution with unknown shape and scale parameters α_i and β_i, respectively, $i = 1, \ldots, k$. That is, G_i has a density function $g_i(\theta|\alpha_i, \beta_i)$, where

$$g_i(\theta|\alpha_i, \beta_i) = \beta_i^{\alpha_i} \theta^{\alpha_i - 1} e^{-\beta_i \theta}/\Gamma(\alpha_i), \quad \theta > 0.$$

Then, $X_{i1}, \ldots, X_{in}$ are iid with marginal probability function $f_i(x) = \Gamma(x + \alpha_i)\beta_i^{\alpha_i}/$ $[\Gamma(\alpha_i)(1 + \beta_i)^{x+\alpha_i}x!]$, $x = 0, 1, 2, \ldots$. Also, $\varphi_i(x) = \frac{x+\alpha_i}{1+\beta_i}$. Straightforward computations yield that $\mu_{i1} \equiv E[X_{i1}] = \alpha_i/\beta_i$, $\mu_{i2} \equiv E[X_{i1}^2] = \alpha_i(\alpha_i + 1)\beta_i^{-2} + \alpha_i\beta_i^{-1}$. Thus, $\beta_i = \mu_{i1}(\mu_{i2} - \mu_{i1} - \mu_{i1}^2)^{-1}$ and $\alpha_i = \mu_{i1}^2(\mu_{i2} - \mu_{i1} - \mu_{i1}^2)^{-1}$. Therefore, $\varphi_i(x) = [x(\mu_{i2} - \mu_{i1} - \mu_{i1}^2) + \mu_{i1}^2](\mu_{i2} - \mu_{i1}^2)^{-1}$.

For each $i = 1, \ldots, k$, let $\mu_{i1n} = n^{-1} \sum_{j=1}^{n} X_{ij}$ and $\mu_{i2n} = n^{-1} \sum_{j=1}^{n} X_{ij}^2$. That is, μ_{i1n} and μ_{i2n} are moment estimators of μ_{i1} and μ_{i2}, respectively. Since it is possible that $\mu_{i2n} - \mu_{i1n} - \mu_{i1n}^2 \equiv \gamma_{in} \leq 0$ though $\mu_{i2} - \mu_{i1} - \mu_{i1}^2 > 0$, thus, for each $x = 0, 1, \ldots$, define

$$\hat{\varphi}_{in}(x) = \begin{cases} \frac{x\gamma_{in} + \mu_{i1n}^2}{\mu_{i2n} - \mu_{i1n}^2} & \text{if } \gamma_{in} > 0, \\ x & \text{otherwise.} \end{cases} \tag{3.6}$$

Then, an empirical Bayes rule $\hat{d}_n = (\hat{d}_{1n}, \ldots, \hat{d}_{kn})$ is proposed as follows: For each $i = 1, \ldots, k$, and $\underline{x} \in \mathcal{X}$,

$$\hat{d}_{in}(\underline{x}) = \begin{cases} 1 & \text{if } \hat{\varphi}_{in}(x_i) \geq \theta_0; \\ 0 & \text{otherwise.} \end{cases} \tag{3.7}$$

3.2.3. A Hierarchical Empirical Bayes Rule

Suppose that the prior distribution G_i is a gamma distribution with a known shape parameter α_i and an unknown scale parameter β_i. In this situation, the preceding parametric empirical Bayes approach can be applied here. However, a new method, called hierarchical empirical Bayes, is introduced in the following.

Since β_i is a scale parameter, it is assumed that β_i has an improper prior $p(\beta_i) = \frac{1}{\beta_i}$, $\beta_i > 0$. Thus, conditional on β_i, $X_{i1}, \ldots, X_{in}$ are iid with the probability function $f_i(x|\beta_i) = \int_0^{\infty} f_i(x|\theta)g_i(\theta|\alpha_i, \beta_i)d\theta = \frac{\Gamma(x+\alpha_i)\beta_i^{\alpha_i}}{x!\Gamma(\alpha_i)(1+\beta_i)^{x+\alpha_i}}$, $x = 0, 1, 2, \ldots$. Therefore, $X_{i1}, \ldots, X_{in}$ has a joint marginal probability function $f_i(x_{i1}, \ldots, x_{in})$, where

$$f_i(x_{i1}, \ldots, x_{in}) = \int_0^{\infty} \prod_{j=1}^{n} f_i(x_{ij}|\beta)p(\beta)d\beta$$

$$= \prod_{j=1}^{n} \left[\frac{\Gamma(x_{ij} + \alpha_i)}{x_{ij}!\Gamma(\alpha_i)} \right] \Gamma(n\alpha_i)\Gamma(b_i - n\alpha_i)/\Gamma(b_i)$$

11

where $b_i = n\alpha_i + \sum_{j=1}^{n} x_{ij}$. Thus, the posterior density function of β_i given $(X_{i1}, \ldots, X_{in}) = (x_{i1}, \ldots, x_{in})$ is

$$p(\beta_i | x_{i1}, \ldots, x_{in}) = \beta_i^{n\alpha_i - 1}(1 + \beta_i)^{-b_i} \Gamma(b_i) / [\Gamma(n\alpha_i)\Gamma(b_i - n\alpha_i)],$$

and the posterior mean of β_i given $(x_{i1}, \ldots, x_{in})$ is

$$\beta_{in} \equiv E[\beta_i | x_{i1}, \ldots, x_{in}] = \begin{cases} \dfrac{n\alpha_i}{\sum_{j=1}^{n} x_{ij} - 1} & \text{if } \sum_{j=1}^{n} x_{ij} \geq 2, \\ \infty & \text{otherwise.} \end{cases}$$

For each $i = 1, \ldots, k$, and $(x_{i1}, \ldots, x_{in})$, define

$$\overline{\varphi}_{in}(x_i) = \begin{cases} \frac{x_i + \alpha_i}{1 + \beta_{in}} & \text{if } \sum_{j=1}^{n} x_{ij} \geq 2; \\ 0 & \text{otherwise.} \end{cases} \tag{3.8}$$

We then propose an empirical Bayes rule $\overline{d}_n = (\overline{d}_{1n}, \ldots, \overline{d}_{kn})$ as follows: For each $i = 1, \ldots, k$,

$$\overline{d}_{in}(x) = \begin{cases} 1 & \text{if } \overline{\varphi}_{in}(x_i) \geq \theta_0; \\ 0 & \text{otherwise.} \end{cases} \tag{3.9}$$

3.2.4. Asymptotic Optimality

For an empirical Bayes selection rule d_n, let $r(G, d_n)$ denote the overall Bayes risk. That is,

$$r(G, d_n) = \sum_{i=1}^{k} \left[\sum_{x_i=0}^{\infty} [\theta_0 - \varphi_i(x_i)] E_{in}[d_{in}(x_i)] f_i(x_i) + C_i \right]$$

where the expectation E_{in} is taken with respect to $(X_{i1}, \ldots, X_{in})$. Since $r(G)$ is the minimal Bayes risk $r(G, d_n) - r(G) \geq 0$ for all n. The nonnegative difference $r(G, d_n) - r(G)$ can be used as a measure of optimality of the empirical Bayes rule d_n.

Definition 3.1. Let $\{d_n\}_{n=1}^{\infty}$ be a sequence of empirical Bayes rules. $\{d_n\}_{n=1}^{\infty}$ is said to be asymptotically optimal of order τ_n relative to the prior distribution G if $r(G, d_n) - r(G) = 0(\tau_n)$, where $\{\tau_n\}_{n=1}^{\infty}$ is a sequence of positive numbers such that $\lim_{n \to \infty} \tau_n = 0$.

Following Gupta and Liang (1989c), it is easy to obtain the following result. Let $B_i(\theta_0) = \{x|\varphi_i(x) < \theta_0\}$ and let

$$m_i = \begin{cases} \max B_i(\theta_0) & \text{if } B_i(\theta_0) \neq \phi; \\ -1 & \text{otherwise} \end{cases}.$$

Theorem 3.1. Let $\underset{\sim}{d}_n$ denote any of the three precedingly constructed empirical Bayes selection rules $\underset{\sim}{d}_n^*, \underset{\sim}{\hat{d}}_n$ and $\underset{\sim}{\bar{d}}_n$. Suppose that $\int_0^\infty \theta dG_i(\theta) < \infty$ and $m_i < \infty$ for all $i = 1, \ldots, k$. Then, $r(\underset{\sim}{G}, \underset{\sim}{d}_n) - r(\underset{\sim}{G}) = O(\exp(-cn))$ for some positive constant c, where the value of c varies depending on the empirical Bayes selection rule used.

3.3. Incorporating Information from Other Components

We now consider the case where it is assumed that the k prior distributions $G_1, \ldots, G_k$ are identical, but there is no past observations available. Under this assumption, $X_1, \ldots, X_k$ are marginally iid with probability function $f(x) = \int_0^\infty e^{-\theta}\theta^x/x! dG(\theta)$ where $G = G_1 = \ldots = G_k$. Therefore, we can still incorporate information from each other to improve the decisions for each of the k component decision problems. The idea is described again through the nonparametric empirical Bayes, the parametric empirical Bayes and the hierarchical empirical Bayes approaches.

3.3.1. A Nonparametric Empirical Bayes Rule

It is assumed that the prior distribution G is completely unknown. Following the discussion of Subsection 3.2.1, a nonparametric empirical Bayes selection rule is constructed as follows.

For each $i = 1, \ldots, k$, let $N_{ik} = \max_{j \neq i} X_j - 1$, and let $f_{ik}(y) = \frac{1}{k-1} \sum_{\substack{j=1 \\ j \neq i}}^k I_{\{y\}}(X_j)$, $h_{ik}(y) = f_{ik}(y)/a(y)$, $y = 0, 1, \ldots$. Also, let $\varphi_{ik}(y) = [h_{ik}(y+1) + \delta_k]/[h_{ik}(y) + \delta_k]$ for each $y = 0, 1, \ldots, N_{ik}$, where $\delta_k > 0$ is such that $\delta_k = o(1)$.

Let $\{\varphi_{ik}^*(y)\}_{y=0}^{N_{ik}}$ be the isotonic regression of $\{\varphi_{ik}(y)\}_{y=0}^{N_{ik}}$ with random weights $\{W_{ik}(y)\}_{y=0}^{N_{ik}}$, where $W_{ik}(y) = [h_{ik}(y) + \delta_k]a(y+1)$. For $y > N_{ik}$, let $\varphi_{ik}^*(y) = \varphi_{ik}^*(N_{ik})$. Now, an empirical Bayes rule $\underset{\sim}{d}_k^* = (d_{1k}^*, \ldots, d_{kk}^*)$ is proposed as follows: For each $i = 1, \ldots, k$, and

13

$(X_1, \ldots, X_k) = (x_1, \ldots, x_k)$, define

$$d_{ik}^*(x) = \begin{cases} 1 & \text{if } \varphi_{ik}^*(x_i) \geq \theta_0; \\ 0 & \text{otherwise.} \end{cases} \tag{3.10}$$

3.3.2. A Parametric Empirical Bayes Rule

It is assumed that the prior distribution G is a member of gamma distribution family with probability density function $g(\theta|\alpha, \beta)$, where

$$g(\theta|\alpha, \beta) = \beta^\alpha \theta^{\alpha-1} e^{-\beta\theta} / \Gamma(\alpha), \quad \theta > 0$$

and both the parameters α and β are unknown. Following the discussion of Subsection 3.2.2, for each $i = 1, \ldots, k$, let $\mu_{1k}(i) = \frac{1}{k-1} \sum_{\substack{j=1 \\ j \neq i}}^{k} X_j$, and $\mu_{2k}(i) = \frac{1}{k-1} \sum_{\substack{j=1 \\ j \neq i}}^{k} X_j^2$. Let $\tau_{ik} = \mu_{2k}(i) - \mu_{1k}(i) - \mu_{1k}^2(i)$. Define

$$\hat{\varphi}_{ik}(x_i) = \begin{cases} \frac{x_i \tau_{ik} + \mu_{1k}^2(i)}{\mu_{2k}(i) - \mu_{1k}^2(i)} & \text{if } \tau_{ik} > 0; \\ x_i & \text{otherwise.} \end{cases} \tag{3.11}$$

An empirical Bayes rule $\hat{d}_k = (\hat{d}_{1k}, \ldots, \hat{d}_{kk})$ is proposed as follows: For each $i = 1, \ldots, k$ and $(X_1, \ldots, X_k) = (x_1, \ldots, x_k)$, define

$$\hat{d}_{ik}(x) = \begin{cases} 1 & \text{if } \hat{\varphi}_{ik}(x_i) \geq \theta_0; \\ 0 & \text{otherwise.} \end{cases} \tag{3.12}$$

3.3.3. A Hierarchical Empirical Bayes Rule

It is assumed that the prior distribution G is a gamma distribution with a known shape parameter α and an unknown scale parameter β. Similar to that of Subsection 3.2.3, a hierarchical empirical Bayes rule $\overline{d}_k = (\overline{d}_{1k}, \ldots, \overline{d}_{kk})$ is constructed as follows.

For given $(X_1, \ldots, X_k) = (x_1, \ldots, x_k)$, let

$$\beta_k = \begin{cases} k\alpha / \left(\sum_{j=1}^{k} x_j - 1 \right) & \text{if } \sum_{j=1}^{k} x_j \geq 2; \\ \infty & \text{otherwise.} \end{cases}$$

14

For each $i = 1, \ldots, k$, and $(X_1, \ldots, X_k) = (x_1, \ldots, x_k)$, define

$$\overline{\varphi}_{ik}(x_i) = \begin{cases} (x_i + \alpha)/(1 + \beta_k) & \text{if } \sum_{j=1}^{k} x_j \geq 2; \\ 0 & \text{otherwise.} \end{cases} \tag{3.13}$$

Define, for each $i = 1, \ldots, k$, and $(X_1, \ldots, X_k) = (x_1, \ldots, x_k)$,

$$\overline{d}_{ik}(x) = \begin{cases} 1 & \text{if } \overline{\varphi}_{ik}(x_i) \geq \theta_0 \\ 0 & \text{otherwise.} \end{cases} \tag{3.14}$$

3.3.4. Asymptotic Optimality

Let $\underline{d}_k$ denote any of the three precedingly constructed empirical Bayes selection rules. The associated overall Bayes risk $r(\underline{G}, \underline{d}_k)$ is:

$$r(\underline{G}, \underline{d}_k) = \sum_{i=1}^{k} r_i(\underline{G}, d_{ik}),$$

where

$$r_i(\underline{G}, d_{ik}) = E_{ik} E_i [(\theta_0 - \varphi_i(X_i)) d_{ik}(\underline{X})] + C$$

where the expectation E_i is taken with respect to X_i and the expectation E_{ik} is taken with respect to $(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_k)$. Also, here $C = \int_{\theta_0}^{\infty} (\theta - \theta_0) dG(\theta)$.

Since $r(\underline{G})$ is the minimal Bayes risk, $r(\underline{G}, \underline{d}_k) - r(\underline{G}) \geq 0$ for all k.

Definition 3.2.

(a) A selection rule $\underline{d}_k$ is said to be weakly asymptotically optimal relative to the prior distribution $\underline{G}$ if

$$[r(\underline{G}, \underline{d}_k) - r(\underline{G})]/k \to 0 \text{ as } k \to \infty.$$

(b) A selection rule $\underline{d}_k$ is said to be strongly asymptotically optimal relative to the prior distribution $\underline{G}$ if

$$r(\underline{G}, \underline{d}_k) - r(\underline{G}) \to 0 \text{ as } k \to \infty.$$

Note that the strong asymptotic optimality implies the weak asymptotic optimality. The weak asymptotic optimality of compound decision rules has been studied in the literature by many authors, notably Vardeman (1978,1980), Gilliland and Hannan (1986), and

Gilliland, Hannan and Hwang (1976), though the formulation of their compound decision problems are different from the one described previously. For the present problem, Gupta and Liang (1989c) obtained the following strong asymptotic optimality.

Let $B(\theta_0) = \{x | \varphi(x) < \theta_0\}$ where $\varphi(x) = \varphi_1(x) = \ldots = \varphi_k(x)$ since $G_1 = \ldots = G_k$ and let

$$m = \begin{cases} \max B(\theta_0) & \text{if } B(\theta_0) \neq \phi, \\ -1 & \text{otherwise.} \end{cases}$$

Theorem 3.2. Let $\underset{\sim}{d}_k$ denote any of the three precedingly constructed empirical Bayes selection rules $\underset{\sim}{d}_k^*, \hat{\underset{\sim}{d}}_k$ and $\overline{\underset{\sim}{d}}$. Suppose that $\int_0^\infty \theta dG(\theta) < \infty$ and $m < \infty$. Then, $r(\underset{\sim}{G}, \underset{\sim}{d}_k) - r(\underset{\sim}{G}) = O(\exp(-ck + \ln k))$ for some positive constant c, where the value of c varies depending on the empirical Bayes rule used.

4. Selection of Variables in Linear Regression

In applying regression analysis in practical situations for prediction purposes such as economic forecasting or weather prediction, one is faced with a large number of independent variables. In such situations, it may well be sufficient to consider only a subset of these predictor variables for an "adequate" prediction. Thus arises a problem of choosing a "good" subset of these variables. Hocking (1976) and Thompson (1978a,b) have reviewed several criteria and techniques that have been used in practice. However, these procedures are ad hoc in nature and are not designed to control the probability of selecting the important variables. McCabe and Arvesen (1974) and Arvesen and McCabe (1975) were first to formulate the problem in the framework of Gupta-type subset selection by considering models involving all possible subsets of an *arbitrarily chosen* size. Huang and Panchapakesan (1982) considered a different formulation taking into consideration all possible reduced models. Using different criteria for comparing any reduced model with the "true" model, this problem was also investigated by Hsu and Huang (1982) who used a sequential procedure, and by Gupta, Huang and Chang (1984) who used simultaneous tests of a family of hypotheses in constructing their procedure. Recently, Gupta and Huang (1988,1989) have further studied this problem. We discuss their results below.

16

Consider the standard linear model

$$\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon} \tag{4.1}$$

where $\underset{\sim}{Y}' = (Y_1, \ldots, Y_n)$ is an n–vector of random observations, $X = [\underset{\sim}{1}, X_1, \ldots, X_{p-1}]$ is an $n \times p$ matrix of known constants, $\underset{\sim}{\beta}' = (\beta_0, \beta_1, \ldots, \beta_{p-1})$ is a p–vector of unknown parameters, and $\underset{\sim}{\varepsilon} \sim N(\underset{\sim}{0}, \sigma^2 I_n)$. Here $\underset{\sim}{1}$ is a column vector of 1's and I_n is an $n \times n$ identity matrix. The model (4.1) with $p-1$ independent variables is considered as the "true" model. Any reduced model whose "X matrix" has r columns is obtained by retaining any $r - 1$ of the $p - 1$ independent variables $X_1, X_2, \ldots, X_{p-1}$, where $2 \leq r \leq p$. For each r, there are $k_r = \binom{p-1}{r-1}$ such models, which are indexed arbitrarily $i = 1, \ldots, k_r$. A typical model from this group will be referred to as M_{ri}, which can be written as

$$E(\underset{\sim}{Y}) = X_{ri}\underset{\sim}{\beta}_{ri} \tag{4.2}$$

where X_{ri} and $\underset{\sim}{\beta}_{ri}$ are obtained from X and $\underset{\sim}{\beta}$, respectively, corresponding to the variables that are retained in the model. In our discussion, *all expectations and probabilities are calculated under the true model (4.1)*.

Let SS_{ri} denote the residual sum of squares for the reduced model $M_{ri}, 1 \leq i \leq k_r, 2 \leq r \leq p$. Then

$$SS_{ri}/\sigma_0^2 \sim \chi^2\{\nu_r, \lambda_{ri}\} \tag{4.3}$$

where $\nu_r = n - r$ is the degrees of freedom and λ_{ri} is the noncentrality parameter. This gives

$$E(SS_{ri}) = \nu_r\sigma_0^2 + 2\sigma_0^2\lambda_{ri}. \tag{4.4}$$

Since σ_0^2 is fixed, it is clear from (4.4) that λ_{ri} should not be large for a good model. This motivates the criterion employed by Gupta and Huang (1988), namely, any reduced model M_{ri} with the associated noncentrality parameter λ_{ri} is defined to be *inferior* if $\lambda_{ri} \geq \Delta$, where $\Delta > 0$ is a specified constant. The goal is to eliminate all inferior models from the set of $2^{p-1} - 1$ regression models including the true model. For this goal, Gupta and Huang (1988) proposed and studied a two–stage procedure. In the first stage, inferior models are

eliminated. Then, in the second stage, one of the models from the retained set (if it has more than one) is selected.

Consider, as an estimator of λ_{ri},

$$\begin{aligned}
\hat{\lambda}_{ri} &= \frac{n-p}{2} \ \frac{SS_{ri}}{SS_{pi}} - \frac{\nu_r}{2} \\
&= \frac{n-p}{2} \ \frac{1-R_{ri}^2}{1-R^2} - \frac{\nu_r}{2}
\end{aligned} \tag{4.5}$$

where R^2 and R_{ri}^2 are the multiple correlation coefficients of the models (4.1) and (4.2), respectively. Define, for $n - p > 2$,

$$\hat{\Gamma}_{ri} = 2 \ \frac{n-p-2}{n-p}[2\hat{\lambda}_{ri} + (p-r)] - (2p - 3r). \tag{4.6}$$

Gupta and Huang (1988) have shown that $\hat{\Gamma}_{ri}$ is an unbiased estimator of $\Gamma_{ri} = \frac{E(SS_{ri})}{\sigma_0^2} - (n - 2r)$, which is the standardized total squared error.

The two–stage procedure R_s of Gupta and Huang (1988) is as follows:

R_S: At stage 1, eliminate all models M_{ri} for which

$$\hat{\lambda}_{ri} \geq d_r \tag{4.7}$$

and at stage 2, select from all the models that are retained after stage 1 the one with the smallest $\hat{\Gamma}_{ri}$. The constant d_r in (4.7) is chosen to satisfy

$$D_r = \left[\left(d_r + \frac{\nu_r}{2}\right) \frac{2}{n-p} - 1\right] \frac{n-p}{n-r} \tag{4.8}$$

where D_r is the $100(1 - P^*)$ percent point of the noncentral F distribution with $p - r$ and $n - p$ degrees of freedom and noncentrality parameter Δ. It can be shown that, for the rule R_s,

$$P\{\text{all inferior models } M_{ri} \text{ are eliminated}\} \geq P^*.$$

Several authors have studied the influence on the fitted regression line when a part of the data is deleted. In the model (4.1), let $\hat{\beta}$ denote the usual least squares estimator of β based on the full data and let $\hat{\beta}_A$ be the least squares estimator based on a subset of the

data. An empirical influence function for $\hat{\beta}$ is $IF_A = \hat{\beta}_A - \hat{\beta}$. For a given positive definite matrix M and a nonzero scale factor c, Cook and Weisberg (1980) defined a distance $D_A(M, c)$ between $\hat{\beta}$ and $\hat{\beta}_A$ given by:

$$D_A(M, c) = \frac{(IF_A)'M(IF_A)}{c}$$

where M can be chosen to reflect specific interests. Recently, Gupta and Huang (1989) have integrated this concept of influential data with their procedure for selecting important independent variables discussed previously. They have considered deleting one observation at a time from the data set Y. Recalling that M_{ri} denotes a reduced model obtained by retaining $r - 1$ of the $p - 1$ independent variables, let $M_{ri(\ell)}$ denote the model obtained from M_{ri} by deleting the ℓ-th observation in Y. Corresponding to λ_{ri} in (4.3) associated with the model M_{ri}, we have the noncentrality parameter $\lambda_{ri(\ell)}$ associated with the model $M_{ri(\ell)}$. Analogous to $\hat{\lambda}_{ri}$ of (4.5) for the model M_{ri}, we define, in the case of $M_{ri(\ell)}$,

$$\hat{\lambda}_{ri(\ell)} = \frac{n - p - 1}{2} \frac{SS_{ri(\ell)}}{SS_{p1(\ell)}} - \frac{n - r - 1}{2}. \tag{4.9}$$

We can find a constant d'_r such that

$$\inf_{\lambda_{ri(\ell)} \geq \Delta} P\{\hat{\lambda}_{ri(\ell)} \geq d'_r\} = P^*. \tag{4.10}$$

The new two–stage procedure R'_s of Gupta and Huang (1989) is defined exactly as their earlier procedure R_s except that, in stage 1, a model M_{ri} is eliminated if

$$\hat{\lambda}_{ri(\ell)} \geq d'_r \text{ for some } \ell \text{ for which } \hat{\lambda}_{p1(\ell)} < d'_p \tag{4.11}$$

instead of (4.7).

5. Sequential Selection Procedures

A substantial amount of original research on sequential selection procedures accomplished during the early years of the ranking and selection theory was published as a monograph by Bechhofer, Kiefer and Sobel (1968). These and subsequent developments

19

have been discussed in Gupta and Panchapakesan (1979), who have recently (1990) reviewed further developments in the sequential selection theory. In our present discussion, we will confine our attention to a few specific recent results.

5.1. A Subset Selection Procedure with a New Goal

Let $\pi_1, \ldots, \pi_k$ be k independent normal populations with unknown means $\theta_1, \ldots, \theta_k$, respectively, and a common *known* variance σ^2. For a specified $\delta^* > 0$, any population π_i for which $\theta_i \geq \max_{1 \leq j \leq k} \theta_j - \delta^*$ is defined to be a *good* population. Gupta and Liang (1988c) considered the goal of selecting a subset of the k populations which *includes the best population* (the one associated with the largest θ_i) *and at the same time excludes all that are not good*. An event of selecting a subset consistent with this goal is denoted by $CS(\delta^*)$. This is different from what is known as δ^*–correct selection in the literature.

Let $X_{i1}, X_{i2}, \ldots$ be a sequence of independent observations from π_i, $i = 1, \ldots, k$. For $m \geq 1$, define $Y_{im} = \sum_{j=1}^{m} X_{ij}$. Let S_m denote the set of contending populations at the beginning of stage m and let $|S_m|$ denote the size of S_m. Gupta and Liang (1988c) proposed and studied the following procedure.

$R_{N:GL}$: Choose a δ_1 in $(0, \delta^*/2)$. At stage $m(m = 1, 2, \ldots)$, take one observation from each population in S_m. *Include* in S_{m+1} *only* those π_i's in S_m for which

$$\frac{\delta_1}{2}(Y_{rm} - Y_{im}) - \frac{m\delta_1^2}{4} < \log \frac{k-1}{1-p^*} \text{ for all } \pi_r \in S_m, \ r \neq i,$$

and *eliminate* all other π_i's from any further consideration. Now, label as good *only* those π_i's in S_{m+1} *that have not been labeled yet* and for which

$$\frac{\delta_1 + \delta^*}{2}(Y_{im} - Y_{tm}) + \frac{m(\delta^{*2} - \delta_1^2)}{4} \geq \log \frac{k-1}{1-p^*} \text{ for all } \pi_i \in S_{m+1}, \ t \neq i.$$

Stop sampling if either $|S_{m+1}| = 1$ or S_{m+1} does not contain any unlabeled population, and *make the terminal decision:* "Select all populations in S_{m+1}", otherwise, go to stage $m + 1$.

It should be noted that a population is not labeled until and unless it qualifies to be called good. Any population, once labeled, is not examined for labeling again. However,

20

it is possible that a labeled population is eliminated subsequently. The populations that are selected by the terminal decision are precisely those which have been found good at some stage and which have survived elimination. The choice of δ_1 in $(0, \delta^*/2)$ assures that the sequential procedure terminates with probability one. The procedure guarantees a minimum probability P^* for selecting a subset consistent with the goal. An optimal choice of δ_1, however, is an open question.

Finally, it should be pointed out that Gupta and Liang (1988c) have discussed the procedure more generally for location and scale parameter families.

5.2. Selection Procedures for the Exponential Family

Gupta and Miescke (1984) studied sequential selection for exponential family under a decision–theoretic framework. Their treatment includes multi–stage selection and their results relate to selection of subsets of random as well as fixed sizes.

Consider the one–parameter exponential family $\mathcal{F}$ given by

$$\mathcal{F} = \{c(\theta) \exp(\theta x) h(x), \ x \in R\}_{\theta \in \Theta}$$

where $\Theta \subseteq R$ is an interval. We consider the class of permutation invariant sequential procedures with or without elimination, employing vector–at–a–time sampling, which means that a vector of observations (one from each) is taken from the non–eliminated populations. Let $X_{i1}, X_{i2}, \ldots$ be a sequence of observations from π_i (with associated parameter θ_i). At stage m ($m = 1, 2, \ldots$), let n_m observations be taken from eligible populations. Let $W_{im} = \sum_{j=1}^{N_m} X_{ij}$, where $N_m = \sum_{j=1}^{m} n_j$, be a sufficient statistic for θ_i, based on all observations from π_i through stage m, and let $\underline{W}_m = (W_{1m}, \ldots, W_{km})$, $m = 1, 2, \ldots$.

Let t_j, $j = 1, \ldots, m$, denote the subset of $\{\pi_1, \ldots, \pi_k\}$ that is eliminated at stage j, and t_{m+1} denote the subset finally selected at termination. This yields a partition $\{t_1, \ldots, t_m, t_{m+1}\}$ of $\{\pi_1, \ldots, \pi_k\}$ which will be called a *record*. For $\underline{\theta} = (\theta_1, \ldots, \theta_k) \in \Omega = \Theta^k$, $L_m(\underline{\theta}; t_1, \ldots, t_m, t_{m+1})$ denotes the loss incurred when the procedure stops at stage m with the record $\{t_1, \ldots, t_m, t_{m+1}\}$. It is assumed that (a) L_m is permutation invariant,

and (b) L_m increases if a record is changed so that a better population is eliminated before an inferior one.

A natural terminal decision, at stage m, selects only those populations among the noneliminated ones which yielded the largest values of W_{im}. Gupta and Miescke (1984) have shown that between two procedures which differ *only* in their terminal decisions, the procedure that employs a natural terminal decision rule has a smaller risk.

One can naturally speculate that, within stages where a procedure with elimination does not stop, natural subset selections are optimal as in the case of terminal decisions. This has been shown to be true by Gupta and Miescke (1984) only in the case of multi-stage procedures with sizes of the subsets selected at each stage fixed, under the assumption that $\mathcal{F}$ is *strongly unimodal* (i.e. exponential density is logconcave). For additional comments, see Miescke (1984).

For the exponential family, Liang (1988) considered the goal of selecting the best population and excluding all that are not good (same goal as that of $R_{N:GL}$ discussed in Section 5.1). His sequential procedure with elimination is based on certain conditional likelihood functions and it achieves the P^*-requirement for $CS(\delta^*)$.

5.3. Other Developments

There are other recent developments concerning, among other things, truncated versions of earlier open sequential procedures, improvements in Paulson's (1964) procedure, and two–factor model with no interaction. For a discussion of these and other developments, see Gupta and Panchapakesan (1990).

6. Lower Confidence Bounds for the Probability of a Correct Selection

Let $X_{ij}, j = 1, \ldots, n$, be a sample of size n from a population π_i, where $\pi_1, \ldots, \pi_k$ are independently distributed with continuous distribution function $G(x - \theta_i)$, $1 \leq i \leq k$. Let $\theta_{[1]} \leq \ldots \leq \theta_{[k]}$ denote the ordered θ_i. The population associated with $\theta_{[k]}$ is called the best population. Assume that the experimenter is interested in the selection of the best population. For this purpose, an appropriate statistic $Y_i = Y(X_{i1}, \ldots, X_{in})$

with cumulative distribution function $F_n(y - \theta_i)$ is chosen, and the natural selection rule that selects the population yielding the largest Y_i as the best population is applied. Let CS (correct selection) denote the event that the best population is selected. Then, the probability of a correct selection (PCS) applying the natural selection rule is: For $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_k)$,

$$P_{\underset{\sim}{\theta}}\{CS\} = \int_{-\infty}^{\infty} \prod_{i=1}^{k-1} F_n(y + \theta_{[k]} - \theta_{[i]}) dF_n(y). \qquad (6.1)$$

To guarantee the PCS, Bechhofer (1954) introduced the indifference zone approach in which the experimenter is asked to assign a positive value δ^* such that $\theta_{[k]} - \theta_{[k-1]} \geq \delta^*$. However, in a real situation, it may be hard to assign the value of δ^* such that $\theta_{[k]} - \theta_{[k-1]} \geq \delta^*$, since the parameter values $\theta_1, \ldots, \theta_k$ are unknown. So that if the above assumption is not satisfied, the PCS cannot be guaranteed to be at least equal to the prespecified level. Parnes and Srinivasan (1986) have pointed out certain inconsistencies in the indifference zone formulation of certain selection problems. Also, see Fabian (1962) and Hsu (1981) for some possible ways to be out of this impasse.

Retrospective analyses regarding the PCS have been studied by several authors. Olkin, Sobel and Tong (1976,1982) have presented estimators of the PCS. Faltin and McCulloch (1983) have studied the small–sample properties of the Olkin–Sobel–Tong estimators for $k = 2$ case. Bofinger (1985) has discussed the nonexistence of consistent estimators of the PCS. Gutmann and Maymin (1987) have presented a procedure to test whether the selected population is the best. Anderson, Bishop and Dudewicz (1977) have given a lower confidence bound on the PCS in normal distribution models.

In the following, we will review some recent developments regarding the construction of lower confidence bounds for the PCS.

6.1. A Lower Confidence Bound on PCS for Distributions with MLR Property

In (6.1), replace $\theta_{[k]} - \theta_{[i]}$ by $\theta_{[k]} - \theta_{[k-1]}$ for each $i = 1, \ldots, k - 2$. Then, one can obtain an inequality

$$P_{\underset{\sim}{\theta}}\{CS\} \geq \int_{-\infty}^{\infty} [F_n(y + \theta_{[k]} - \theta_{[k-1]})]^{k-1} dF_n(y). \qquad (6.2)$$

Kim (1986) proposed a method to find a conservative lower confidence bound on the PCS by first finding a lower confidence bound on $\theta_{[k]} - \theta_{[k-1]}$.

Let $H(t)$ be the distribution of $(Y_1 - \theta_1) - (Y_2 - \theta_2)$. That is,

$$H(x) = \int_{-\infty}^{\infty} F_n(x+y) dF_n(y). \tag{6.3}$$

The distribution $H(x)$ is independent of the parameters θ_1 and θ_2, and $H(x)$ is symmetric about the point 0. For $0 < \alpha < 1$, let $t_{\alpha/2}$ be the upper $\alpha/2$–quantile of the distribution $H(x)$. By the symmetric property of $H(x)$, $t_{\alpha/2} > 0$. For this fixed α, define a nonnegative function $L_\alpha(t)$ on $[0, \infty)$ implicitly by

$$H(L_\alpha(t) - t) + H(-L_\alpha(t) - t) = \alpha \qquad \text{for} \qquad t \geq t_{\alpha/2} \tag{6.4}$$

and let $L_\alpha(t) = 0$ if $0 \leq t < t_{\alpha/2}$. Let $Y_{[1]} \leq \ldots \leq Y_{[k]}$ denote the ordered statistics of $Y_1, \ldots, Y_k$. Also, let f_n be the associated pdf of the distribution function F_n. Finally, define

$$\hat{P}_L = \int_{-\infty}^{\infty} [F_n(y + L_\alpha(Y_{[k]} - Y_{[k-1]}))]^{k-1} dF_n(y). \tag{6.5}$$

Theorem 6.1 (Kim (1986)). Assume that $\log f_n(y)$ is concave. Then,

$$\inf_{\underset{\sim}{\theta}} P_{\underset{\sim}{\theta}}\{\theta_{[k]} - \theta_{[k-1]} \geq L_\alpha(Y_{[k]} - Y_{[k-1]})\} = 1 - \alpha,$$

and hence,

$$P_{\underset{\sim}{\theta}}\{P_{\underset{\sim}{\theta}}\{CS\} \geq \hat{P}_L\} \geq 1 - \alpha \text{ for all } \underset{\sim}{\theta}.$$

6.1.1. Normal Populations with a Common Variance

Let $X_{ij}, j = 1, \ldots, n$, be a sample of size n from $N(\theta_i, \sigma^2)$, $i = 1, \ldots, k$, where the common variance σ^2 may be either known or unknown. The best population is the one associated with $\theta_{[k]}$. Let $Y_i = \frac{1}{n} \sum_{j=1}^{n} X_{ij}$ be the sample mean for each $i = 1, \ldots, k$. The natural selection rule selects the population yielding the largest sample mean value $Y_{[k]}$ as the best population. The PCS is:

$$P_{\underset{\sim}{\theta}}\{CS\} = \int_{-\infty}^{\infty} \prod_{i=1}^{k-1} \Phi\left(x + \frac{\sqrt{n}(\theta_{[k]} - \theta_{[i]})}{\sigma}\right) d\Phi(x),$$

24

where $\Phi(\cdot)$ is the standard normal distribution.

When the common variance σ^2 is known, for $0 < \alpha < 1$, the function $L_\alpha(t)$ is implicitly defined such that

$$\Phi(L_\alpha(t) - t) + \Phi(-L_\alpha(t) - t) = \alpha \qquad \text{for} \qquad t \geq z_{\alpha/2},$$

and $L_\alpha(t) = 0$ for $0 \leq t < z_{\alpha/2}$, where $z_{\alpha/2}$ is the upper $\alpha/2$–quantile of $\Phi(\cdot)$. Kim (1986) obtained a lower confidence bound for the PCS which is given as follows:

$$\hat{P}_L = \int_{-\infty}^{\infty} \left[\Phi\left(x + \sqrt{2}L_\alpha\left(\frac{\sqrt{n}(Y_{[k]} - Y_{[k-1]})}{\sqrt{2}\sigma} \right) \right) \right]^{k-1} d\Phi(x),$$

and $P_\theta\{P_\theta\{CS\} \geq \hat{P}_L\} \geq 1 - \alpha$ for all $\underset{\sim}{\theta}$.

When the common variance σ^2 is unknown, let $S^2 = \frac{1}{\nu} \sum_{i=1}^{k} \sum_{j=1}^{n} (X_{ij} - Y_i)^2$, where $\nu = k(n-1)$. Note that $\frac{\nu S^2}{\sigma^2}$ has a χ^2–distribution with ν degrees of freedom. Let Q_ν denote the distribution of S/σ. For given $0 < \alpha < 1$, let $L_\alpha^*(t)$ be the function implicitly defined by

$$\int_0^{\infty} [\Phi(L_\alpha^*(t) - tu) + \Phi(-L_\alpha^*(t) - tu)]dQ_\nu(u) = \alpha \qquad \text{for} \qquad t \geq t_{\alpha/2}(\nu)$$

and $L_\alpha^*(t) = 0$ for $0 \leq t < t_{\alpha/2}(\nu)$, where $t_{\alpha/2}(\nu)$ is the upper $\alpha/2$–quantile of the t–distribution with ν degrees of freedom. Kim (1986) obtained a lower confidence bound for the PCS as follows:

$$P_L^* = \int_{-\infty}^{\infty} \left[\Phi\left(x + \sqrt{2}L_\alpha^*\left(\frac{\sqrt{n}(Y_{[k]} - Y_{[k-1]})}{\sqrt{2}S} \right) \right) \right]^{k-1} d\Phi(x),$$

and $P_\theta\{P_\theta\{CS\} \geq P_L^*\} \geq 1 - \alpha$ for all $\underset{\sim}{\theta}$.

The table used to implement the procedures has been tabulated by Kim (1986) for $\alpha = 0.5$ and 0.1 for some ν values.

6.1.2. Two–Parameter Exponential Populations

Let $X_{ij}, j = 1, \ldots, n$, be a sample of size n from a two–parameter exponential distribution with pdf $g(x|\theta_i, \beta) = \beta^{-1} \exp(-(x - \theta_i)/\beta)I_{(\theta_i, \infty)}(x)$, $i = 1, \ldots, k$, where the

common scale parameter $\beta > 0$ may be either known or unknown. The best population is the one associated with $\theta_{[k]}$. Let $Y_i = \min(X_{i1}, \ldots, X_{in})$, $i = 1, \ldots, k$. The natural selection rule selects the population yielding $Y_{[k]}$ as the best population. The PCS is:

$$P_{\underset{\sim}{\theta}}\{CS\} = \int_{y=0}^{\infty} \prod_{i=1}^{k-1} \left[1 - \exp(-(y + n(\theta_{[k]} - \theta_{[i]})/\beta))\right] e^{-y} dy.$$

Let $H(t)$ be the distribution function of $\frac{n(Y_1 - \theta_1)}{\beta} - \frac{n(Y_2 - \theta_2)}{\beta}$. Then,

$$H(t) = \begin{cases} 1 - \frac{1}{2}e^{-t} & \text{if } t \geq 0; \\ \frac{1}{2}e^{t} & \text{if } t < 0. \end{cases}$$

When the common scale parameter β is known, for $0 < \alpha < 1$, let $t_{\alpha/2}$ denote the upper $\alpha/2$-quantile of H(t). Then, the function $L_\alpha(t)$ is implicitly defined by

$$H(L_\alpha(t) - t) + H(-L_\alpha(t) - t) = \alpha \text{ for } t \geq t_{\alpha/2}$$

and $L_\alpha(t) = 0$ for $0 \leq t < t_{\alpha/2}$.

Gupta, Leu and Liang (1990) obtained a lower confidence bound for the PCS as follows:

$$\hat{P}_L = \int_{y=0}^{\infty} [1 - \exp(-y - L_\alpha(n(Y_{[k]} - Y_{[k-1]})/\beta))]^{k-1} e^{-y} dy,$$

and $P_{\underset{\sim}{\theta}}\{P_{\underset{\sim}{\theta}}\{CS\} \geq \hat{P}_L\} \geq 1 - \alpha$ for all $\underset{\sim}{\theta}$.

When the common scale parameter β is unknown, let $S = \frac{1}{\nu} \sum_{i=1}^{k} \sum_{j=1}^{n} (X_{ij} - Y_i)$, where $\nu = k(n - 1)$. Then $\frac{\nu S}{\beta}$ has a $\Gamma(\nu, 1)$ distribution. Let $Q_\nu(\cdot)$ denote the distribution of S/β. For $0 < \alpha < 1$, let $t_{\alpha/2}^*$ be the point such that $\int_0^\infty H(-t_{\alpha/2}^* y) dQ_\nu(y) = \alpha/2$. The function $L_\alpha^*(t)$ is then implicitly defined by

$$\int_0^\infty [H(L_\alpha^*(t) - yt) + H(-L_\alpha^*(t) - yt)] dQ_\nu(y) = \alpha \qquad \text{for} \qquad t \geq t_{\alpha/2}^*,$$

and $L_\alpha^*(t) = 0$ for $0 \leq t < t_{\alpha/2}^*$.

Gupta, Leu and Liang (1990) obtained the following lower confidence bound for the PCS:

$$P_L^* = \int_0^\infty [1 - \exp(-y - L_\alpha^*(n(Y_{[k]} - Y_{[k-1]})/S))]^{k-1} e^{-y} dy,$$

and $P_{\underset{\sim}{\theta}}\{P_{\underset{\sim}{\theta}}\{CS\} \geq P_L^*\} \geq 1 - \alpha$ for all $\underset{\sim}{\theta}$.

The table used to implement the procedures has been tabulated by Gupta, Leu and Liang (1990) for $\alpha = 0.05$ and 0.1 for some ν values.

6.2. Lower Confidence Bounds on the PCS for General Location–Parameter Models

Gupta and Liang (1987) have constructed lower confidence bounds on the PCS for general location–parameter models, where the sample size n is determined according to the indifference zone formulation of Bechhofer (1954). Note that

$$\inf_{\underset{\sim}{\theta} \in \Omega_{(\delta^*)}} P_{\underset{\sim}{\theta}}\{CS\} = \int_{-\infty}^{\infty} [F_n(y + \delta^*)]^{k-1} dF_n(y), \tag{6.6}$$

where $\Omega(\delta^*) = \{\underset{\sim}{\theta} | \theta_{[k]} - \theta_{[k-1]} \geq \delta^*\}$ is called the preference zone. Suppose that the right–hand–side of (6.6) is an increasing function of n, and tends to one as n tends to infinity. For a given probability level $P^*(k^{-1} < P^* < 1)$, let

$$n_0 \equiv n_0(\delta^*, P^*) = \min \left\{ n : \int_{-\infty}^{\infty} [F_n(y + \delta^*)]^{k-1} dF_n(y) \geq P^* \right\}. \tag{6.7}$$

That is, n_0 is the minimum common sample size so that the PCS will be guaranteed at least to be P^* when the natural selection rule is applied and $\underset{\sim}{\theta} \in \Omega(\delta^*)$.

Let $Y_{[1]} \leq \ldots \leq Y_{[k]}$ denote the ordered statistics of $Y_1, \ldots, Y_k$. For given $0 < \alpha < 1$, let $c(k, n_0, \alpha)$ be the value such that

$$P_{\underset{\sim}{\theta}} \left\{ \max_{1 \leq i \leq k} (Y_i - \theta_i) - \min_{1 \leq j \leq k} (Y_j - \theta_j) \leq c(k, n_0, \alpha) \right\} = 1 - \alpha. \tag{6.8}$$

Note that the value of $c(k, n_0, \alpha)$ is independent of $\underset{\sim}{\theta}$. Define

$$\tilde{\delta}_{L_i} = (Y_{[k]} - Y_{[i]} - c(k, n_0, \alpha))^+, \tag{6.9}$$

where $y^+ = \max(0, y)$, and

$$\hat{P}_L = \int_{-\infty}^{\infty} \prod_{i=1}^{k-1} F_n(y + \hat{\delta}_{L_i}) dF_n(y). \tag{6.10}$$

27

Gupta and Liang (1987) proposed $\hat{P}_L$ as an estimator of a lower bound of the PCS, and obtained the following result.

Theorem 6.2 (Gupta and Liang (1987)).

$$P_{\underline{\theta}}\{\theta_{[k]} - \theta_{[i]} \geq \hat{\delta}_{L_i} \text{ for all } i = 1, \ldots, k-1\} \geq 1 - \alpha \text{ for all } \underline{\theta},$$

and therefore,

$$P_{\underline{\theta}}\{P_{\underline{\theta}}\{CS\} \geq \hat{P}_L\} \geq 1 - \alpha \text{ for all } \underline{\theta}.$$

6.2.1. Normal Population with a Common Variance

Consider k normal populations $N(\theta_i, \sigma^2), i = 1, \ldots, k$, with unknown means $\theta_1, \ldots, \theta_k$ and common variance σ^2, where σ^2 may be either known or unknown.

When the common variance σ^2 is known, let $Y_i = \frac{1}{n_0} \sum_{j=1}^{n_0} X_{ij}$, where $X_{i1}, \ldots, X_{in_0}$ is a sample of size n_0 from $N(\theta_i, \sigma^2)$ and n_0 is determined, for the indifference zone function, by

$$n_0 = \min\left\{n: \int_{-\infty}^{\infty} \left[\Phi\left(x + \frac{\sqrt{n}\delta^*}{\sigma}\right)\right]^{k-1} d\Phi(x) \geq P^*\right\}$$

where both $\delta^*(> 0)$ and $P^*(k^{-1} < P^* < 1)$ are prespecified by the experimenter. The PCS applying the natural selection rule is

$$P_{\underline{\theta}}\{CS\} = \int_{-\infty}^{\infty} \prod_{i=1}^{k-1} \Phi\left(x + \frac{\sqrt{n_0}(\theta_{[k]} - \theta_{[i]})}{\sigma}\right) d\Phi(x).$$

For given $0 < \alpha < 1$, choose the value $c(k, n_0, \alpha)$ such that

$$P_{\underline{\theta}}\left\{\max_{1 \leq i \leq k}(Y_i - \theta_i) - \min_{1 \leq j \leq k}(Y_j - \theta_j) \leq c(k, n_0, \alpha)\right\} = 1 - \alpha.$$

Note that here, $c(k, n_0, \alpha) = \frac{\sigma}{\sqrt{n_0}} q_{k,\infty}^\alpha$, where $q_{k,\infty}^\alpha$ is the $100(1-\alpha)$%th percentile of Tukey's studentized range statistic with parameter (k, ∞). The value of $q_{k,\infty}^\alpha$ is available from Harter (1965). Define

$$\hat{\delta}_{L_i} = (Y_{[k]} - Y_{[i]} - c(k, n_0, \alpha))^+$$

and

$$\hat{P}_L = \int_{-\infty}^{\infty} \prod_{i=1}^{k-1} \Phi\left(x + \frac{\sqrt{n_0}\hat{\delta}_{L_i}}{\sigma}\right) d\Phi(x).$$

Then, by Theorem 6.2, $P_{\underline{\theta}}\{P_{\underline{\theta}}\{CS\} \geq \hat{P}_L\} \geq 1 - \alpha$ for all $\underline{\theta}$.

When the common variance σ^2 is unknown, Bechhofer, Dunnett and Sobel (1954) presented a two–stage selection rule given as follows.

Take a sample of size $n_0(n_0 \geq 2)$ observations from each of the k populations. Compute $Y_i(n_0) = \frac{1}{n_0} \sum_{j=1}^{n_0} X_{ij}$, $i = 1, \ldots, k$, and $S^2 = \frac{1}{\nu} \sum_{i=1}^{k} \sum_{j=1}^{n_0} (X_{ij} - Y_i(n_0))$, where $\nu = k(n_0 - 1)$. Define $N = \max\left\{n_0, \left[\frac{S^2 h^2}{\delta^{*2}}\right]\right\}$, where $[y]$ is the smallest integer not less than y, and h is a positive value such that

$$\int_0^{\infty} \int_{-\infty}^{\infty} [\Phi(x + wh)]^{k-1} d\Phi(x) dF_W(w) = P^*$$

where $F_W(\cdot)$ is the distribution function of the nonnegative random variable W with νW^2 following a $\chi^2(\nu)$–distribution.

Then, take $N - n_0$ observations from each of the k populations. Compute the overall sample mean $Y_i(N) = \frac{1}{N} \sum_{j=1}^{N} X_{ij}$, $i = 1, \ldots, k$. The natural selection rule selects the population yielding the largest sample mean value $Y_{[k]}(N)$ as the best population.

For this two–stage selection rule,

$$P_{\underline{\theta}}\{CS\} \geq \int_0^{\infty} \int_{-\infty}^{\infty} \prod_{i=1}^{k-1} \Phi\left(x + \frac{h(\theta_{[k]} - \theta_{[i]})w}{\delta^*}\right) d\Phi(x) dF_W(w).$$

Let $c = S q_{k,\nu}^{\alpha}/\sqrt{N}$, where $q_{k,\nu}^{\alpha}$ is the $100(1 - \alpha)\%$th percentile of Tukey's studentized range statistic with parameters (k, ν). Define $\hat{\delta}_{L_i} = (\overline{Y}_{[k]}(N) - \overline{Y}_{[i]}(N) - c)^+$. Let

$$\hat{Q}_L = \int_0^{\infty} \int_{-\infty}^{\infty} \prod_{i=1}^{k-1} \Phi\left(x + \frac{h\hat{\delta}_{L_i}w}{\delta^*}\right) d\Phi(x) dF_W(w).$$

Gupta and Liang (1987) obtained the following lower confidence bound on the PCS:

$$P_{\underline{\theta}}\{\theta_{[k]} - \theta_{[i]} \geq \hat{\delta}_{L_i} \text{ for all } i = 1, \ldots, k - 1\} \geq 1 - \alpha \text{ for all } \underline{\theta},$$

and therefore,

$$P_{\underline{\theta}}\{P_{\underline{\theta}}\{CS\} \geq \hat{Q}_L\} \geq 1 - \alpha \text{ for all } \underline{\theta}.$$

REFERENCES

Anderson, P.O., Bishop, T.A. and Dudewicz, E.J. (1977). Indifference zone ranking and selection: confidence intervals for true achieved P(CD). *Communications in Statistics - Theory and Methods* A(6), 1121-1132.

Arvesen, J.N. and McCabe, G.P., Jr. (1975). Subset selection problems of variances with applications to regression analysis. *Journal of the American Statistical Association* 70, 166-170.

Bechhofer, R.E. (1954). A single-sample multiple-decision procedure for ranking means of normal populations with known variances. *Annals of Mathematical Statistics* 25, 16-39.

Bechhofer, R.E. and Dunnett, C.W. and Sobel, M. (1954). A two-sample multiple decision procedure for ranking means of normal populations with a common unknown variance. *Biometrika* 41, 170-176.

Bechhofer, R.E., Kiefer, J. and Sobel, M. (1968). *Sequential Identification and Ranking Procedures (with special reference to Koopman-Darmois populations)*. The University of Chicago Press, Chicago.

Bofinger, E. (1985). On the non-existence of consistent estimators for $P(CS)$. *American Journal of Mathematics and Management Sciences* 5, 63-76.

Cohen, A. and Sackrowitz, H.B. (1982). Estimating the mean of the selected population. *Statistical Decision Theory and Related Topics - III*, Vol. 1, (Eds. S.S. Gupta and J.O. Berger), Academic Press, New York, 243-270.

Cohen, A. and Sackrowitz, H.B. (1988). A decision theory formulation for population selection followed by estimating the mean of the selected population. *Statistical Decision Theory and Related Topics - IV*, Vol. 2, (Eds. S.S. Gupta and J.O. Berger), Springer Verlag, New York, 33-36.

Cook, R.D. and Weisberg, S. (1980). Characterizations of an empirical influence function for detecting influential cases in regression. *Technometrics* 22, 495-508.

Dahiya, R.C. (1974). Estimation of the mean of the selected population. *Journal of the American Statistical Association* **69**, 226–230.

Deely, J.J. (1965). Multiple decision procedures from an empirical Bayes approach. Ph.D. Thesis (Mimeo Ser. No. 45), Department of Statistics, Purdue University, West Lafayette, Indiana.

Fabian, V. (1962). On multiple decision methods for ranking population means. *Annals of Mathematical Statistics* **33**, 248–254.

Faltin, F. and McCulloch, C.E. (1983). On the small–sample properties of the Olkin–Sobel–Tong estimator of the probability of correct selection. *Journal of the American Statistical Association* **78**, 464–467.

Gilliland, D.C. and Hannan, J. (1986). The finite state compound decision problem, equivalence and restricted risk components. *Adaptive Statistical Procedures and Related Topics* (Ed. J. Van Ryzin), IMS Lecture Notes–Monograph Series, Vol. 8, 129–145.

Gilliland, D.C., Hannan, J. and Huang, J.S. (1976). Asymptotic solutions to the two state component compound decision problem, Bayes versus diffuse priors on proportions. *Ann. Statist.* **4**, 1101–1112.

Gupta, S.S. (1956). On a decision rule for a problem in ranking means. Mimeograph Series No. 150, Institute of Statistics, University of North Carolina, Chapel Hill, North Carolina.

Gupta, S.S. (1965). On some multiple decision (selection and ranking) rules, *Technometrics* **7**, 225–245.

Gupta, S.S. and Hsiao, P. (1983). Empirical Bayes rules for selecting good populations. *J. Statist. Plann. Inference* **8**, 87–101.

Gupta, S.S. and Huang, D.-Y. (1988). Selecting Important Independent Variables in Linear Regression Models. *Journal of Statistical Planning and Inference* **20**, 155–167.

Gupta, S.S. and Huang, D.-Y. (1989). On detecting influential data and selecting regression variables. Tech. Report #89-28C, Department of Statistics, Purdue University, West Lafayette, Indiana.

Gupta, S.S., Huang, D.-Y., and Chang, C.-L. (1984). Selection procedures for optimal subsets of regression variables. *Design of Experiments: Ranking and Selection*, (T.J. Santner and A.C. Tamhane, eds.), Marcel Dekker, New York, 67–75.

Gupta, S.S. and Leu, L.Y. (1983). On Bayes and empirical Bayes rules for selecting good populations. Technical Report #83-37, Department of Statistics, Purdue University, West Lafayette, Indiana.

Gupta, S.S., Leu, L.Y. and Liang, T. (1990). On lower confidence bounds for PCS in truncated location parameter models. *Commun. Statist.-Theory Meth.* 19(2), 527–546.

Gupta, S.S. and Liang, T. (1986). Empirical Bayes rules for selecting good binomial populations. *Adaptive Statistical Procedures and Related Topics* (Ed. J. Van Ryzin), IMS Lecture Notes–Monograph Series, Vol. 8, 110–128.

Gupta, S.S. and Liang, T. (1987). On a lower confidence bound for the probability of a correct selection: analytical and simulation studies. To appear in *The Proceedings of the First International Conference on Statistical Computing*.

Gupta, S.S. and Liang, T. (1988a). Empirical Bayes rules for selecting the best binomial population. *Statistical Decision Theory and Related Topics-IV* (Eds. S.S. Gupta and J.O. Berger), Springer–Verlag, Vol. I, 213–224.

Gupta, S.S. and Liang, T. (1988b). On empirical Bayes selection rules for negative binomial populations. To appear in *The Proceedings of the Indo–US Workshop on Bayesian Analysis in Statistics and Econometrics*.

Gupta, S.S. and Liang, T. (1988c). On a sequential subset selection procedure, Technical Report #88-23, Department of Statistics, Purdue University, West Lafayette, Indiana.

Gupta, S.S. and Liang, T. (1989a). Parametric empirical Bayes rules for selecting the most probable multinomial event. Chapter 18 in *Contributions to Probability and Statistics: Essays in Honor of Ingram Olkin*, Springer–Verlag, 318–328.

Gupta, S.S. and Liang, T. (1989b). Selecting the best binomial population: parametric empirical Bayes approach. *J. Statist. Plann. Inference* **23**, 21–31.

Gupta, S.S. and Liang, T. (1989c). On the asymptotic optimality of certain empirical Bayes simultaneous testing procedures. Technical Report #89–22C, Department of Statistics, Purdue University, West Lafayette, Indiana.

Gupta, S.S. and Miescke, K.J. (1984). Sequential selection procedures – A decision–theoretic approach, *Ann. Statist.* **12**, 336–350.

Gupta, S.S. and Miescke, K.J. (1988). On the problem of finding the largest normal mean under heteroscedasticity. *Statistical Decision Theory and Related Topics - IV*, Vol. **2** (Eds. S.S. Gupta and J.O. Berger), Springer–Verlag, New York, 37–49.

Gupta, S.S. and Miescke, K.J. (1990). On finding the largest normal mean and estimating the selected mean. To appear in *Sankhya B*.

Gupta, S.S. and Panchapakesan, S. (1979). *Multiple Decision Procedures*. Wiley, New York.

Gupta, S.S. and Panchapakesan, S. (1985). Subset selection procedures: review and assessment. *Amer. J. Math. Management Sci.* **5**, 235–311.

Gupta, S.S. and Panchapakesan, S. (1990). On sequential ranking and selection procedures. *Handbook of Sequential Methods* (Eds. B.K. Ghosh and P.K. Sen), Marcel Dekker, New York, to appear.

Gutmann, S. and Maymin, Z. (1987). Is the selected population the best? *Ann. Statist.* **15**, 456–461.

Harter, H.L. (1969). Order statistics and their use in testing and estimation, Vol. 1, Tests based on range and studentized range of samples from a normal population, Aerospace Research Laboratories.

Hocking, R.R. (1976). The analysis and selection of variables in regression analysis, *Biometrics* **32**, 1–49.

Hsu, J.C. (1981). Simultaneous confidence intervals for all distances from the "best". *Annals of Statistics* **9**, 1026–1034.

Hsu, T.-A. and Huang, D.-Y. (1982). Some sequential selection procedures for good regression models. *Communications in Statistics – Theory and Methods* **A11**, 411–421.

Huang, D.-Y. and Panchapakesan, S. (1982). On eliminating inferior regression models. *Communications in Statistics – Theory and Methods* **A11**, 751–759.

Jeyaratnam, S. and Panchapakesan, S. (1984). An estimation problem relating to subset selection for normal populations. *Design of Experiments; Ranking and Selection* (Eds. T.J. Santner and A.C. Tamhane), Marcel Dekker, New York, 287–302.

Jeyaratnam, S. and Panchapakesan, S. (1986). Estimation after subset selection from exponential populations. *Communications in Statistics – Theory and Methods* **10**, 1869–1878.

Jeyaratnam, S. and Panchapakesan, S. (1988). Selections from uniform populations based on sample midranges and an associated estimation after selection. *Communications in Statistics – Theory and Methods* **17**, 2303–2314.

Kim, W.-C. (1986). A lower confidence bound on the probability of a correct selection. *Journal of the American Statistical Association*, **81**, 1012–1017.

Liang, T. (1988). On a sequential subset selection procedure for exponential family distributions, Technical Report #88–20C, Department of Statistics, Purdue University, West Lafayette, Indiana.

McCabe, G.P., Jr. and Arvesen, J.N. (1974). Subset selection procedures for regression variables. *Journal of Statistical Computation and Simulation* **3**, 137–146.

Miescke, K.J. (1984). Recent results on multi–stage selection procedures, Proceedings of the Seventh Conference on Probability Theory (Braşov, 1982), Editura Academiei Republicii Socialiste România, Bucharest, pp. 259–268.

Olkin, I., Sobel, M. and Tong, Y.L. (1976). Estimating the true probability of a correct selection for location and scale parameter families. Technical Report **110**, Stanford University, Department of Statistics.

Olkin, I., Sobel, M. and Tong, Y.L. (1982). Bounds for a k–fold integral for location and scale parameter models with applications to statistical ranking and selection problem. *Statistical Decision Theory and Related Topics III*, Vol. **2** (Eds. S.S. Gupta and J.O. Berger), New York, Academic Press, pp. 193–212.

Parnes, M. and Srinivasan, R. (1986). Some inconsistencies and errors in the indifference zone formulation of selection. *Sankhya* **48**, 86–97.

Paulson, E. (1964). A sequential procedure for selecting the population with the largest mean from k normal populations, *Ann. Math. Statist.* **35**, 174–180.

Robbins, H. (1956). An empirical Bayes approach to statistics. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1**, 157–163, University of California Press.

Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. *Ann. Math. Statist.* **35**, 1–20.

Sackrowitz, H.B. and Samuel–Cahn, E. (1984). Estimation of the mean of a selected negative exponential population. *Journal of the Royal Statistical Society, Series B*, **46**, 242–249.

Sackrowitz, H.B. and Samuel–Cahn, E. (1986). Evaluating the chosen population: A Bayes and minimax approach. *Adaptive Statistical Procedures and Related Topics*, (Ed. J. Van Ryzin), IMS Lecture Notes – Monograph Series, Vol. **8**, 386–399.

Sarkadi, K. (1967). Estimation after selection. *Studia Scientiarum Mathematicarum Hungarica* **2**, 341–350.

Thompson, M.L. (1978a). Selection of variables in multiple regression: Part I. A review and evaluation. *International Statistical Review* **46**, 1–19.

Thompson, M.L. (1978b). Selection of variables in multiple regression: Part II. Chosen procedures, computations and examples. *International Statistical Review* **46**, 129–146.

Vardeman, S.B. (1978). Bounds on the empirical Bayes and compound risks of truncated version of Robbins's estimator of a binomial parameter. *J. Statist. Plann. Inference* **2**, 245–252.

Vardeman, S.B. (1980). Admissible solutions of k–extended finite state set and sequence compound decision problems. *J. Multivariate Anal.* **10**, 426–441.

Vellaisamy, P., Kumar, S., and Sharma, D. (1988). Estimating the mean of the selected uniform population. *Communications in Statistics - Theory and Methods* **17**, 3447–3476.

Vellaisamy, P. and Sharma, D. (1988). Estimating the mean of the selected gamma population. *Communications in Statistics - Theory and Methods* **17**, 2797–2817.

Vellaisamy, P. and Sharma, D. (1989). A note on the estimation of the mean of the selected gamma population. *Communications in Statistics - Theory and Methods* **18**, 555–560.

Venter, J.H. (1988). Estimation of the mean of the selected population. *Communications in Statistics - Theory and Methods* **17**, 791–805.

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release, distribution unlimited.
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	

4. PERFORMING ORGANIZATION REPORT NUMBER(S) Technical Report 90-27C	5. MONITORING ORGANIZATION REPORT NUMBER(S)

6a. NAME OF PERFORMING ORGANIZATION Purdue University	6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION

6c. ADDRESS (City, State, and ZIP Code) Department of Statistics West Lafayette, IN 47907	7b. ADDRESS (City, State, and ZIP Code)

8a. NAME OF FUNDING / SPONSORING ORGANIZATION Office of Naval Research	8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-88-K-0170, NSF Grant DMS-8702620

8c. ADDRESS (City, State, and ZIP Code) Arlington, VA 22217-5000	10. SOURCE OF FUNDING NUMBERS			
	PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)

RECENT ADVANCES IN STATISTICAL RANKING AND SELECTION: THEORY AND METHODS

12. PERSONAL AUTHOR(S)
Shanti S. Gupta

13a. TYPE OF REPORT Technical	13b. TIME COVERED FROM _____ TO _____	14. DATE OF REPORT (Year, Month, Day) June 1990	15. PAGE COUNT

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Ranking and Selection, Empirical Bayes Procedures, Asymptotic Optimality, Probability of Correct Selection, Regression Variables, Sequential Selection Procedures.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Ranking and selection procedures have been developed in modern statistical methodology over the past 35 years with fundamental papers pioneered by Bechhofer (1954) and Gupta (1956). Since then, various modifications and applications have taken place. The reader is referred to Gupta and Panchapakesan (1979, 1985) for an overview of this research area.

In the present paper, we provide a review of some recent developments in the research area of ranking and selection, mainly, on the following topics: (A) Selecting the largest normal mean and estimating the selected mean, (B) empirical Bayes selection, (C) selecting the important regression variables, (D) sequential selection rules, and (E) lower confidence bounds for the probability of a correct selection. Related theoretic and methodological research will be surveyed. Aspects of ongoing research will also be discussed.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Shanti S. Gupta	22b. TELEPHONE (Include Area Code) 317-494-6031 22c. OFFICE SYMBOL

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted. SECURITY CLASSIFICATION OF THIS PAGE
All other editions are obsolete.

UNCLASSIFIED